Infrastructure - Task #3137

Story # 3136 (Closed): Inconsistent data in production 1.0.3 release

Identify and correct missing identifiers table entries on ORC and UNM (w/systemMetadata)

2012-08-15 16:08 - Robert Waltz

Status:	Closed		**Start date:**	2012-08-15
Priority:	Immediate		**Due date:**	
Assignee:	Ben Leinfelder		**% Done:**	100%
Category:	Environment.Production		**Estimated time:**	0.00 hour
Target version:	Sprint-2012.29-Block.4.3			
Milestone:	CCI-1.0.4		**Story Points:**	
Product Version:	*			

Description

The systemMetadata table entries are missing entries in the identifiers table. Any dataONE type that is a Science Metadata or a Resource Map should have a corresponding entry in the identifiers table.

Use the following query on the 3 production machines:

su postgres -c "psql -d metacat -c \"select count(*) from systemMetadata s where \
s.object_format not in ('CF-1.0','CF-1.1','CF-1.2','CF-1.3','CF-1.4','text/plain', \
'text/csv','image/bmp','image/gif','image/jp2','image/jpeg','image/png', \
'image/svg+xml','image/tiff','application/octet-stream','application/pdf') and \
not exists (select * from identifier i where i.guid = s.guid)\""

It will result with 90 missing identiers on ORC, 128 on UNM and 0 on UCSB.

Related issues:

Related to Infrastructure - Task #3697: Missing identifier mappings on all th...	**Closed**	**2013-03-28**

History

#1 - 2012-08-15 18:53 - Ben Leinfelder

- Status changed from New to In Progress

This seems to be a simple Metacat replication issue. Before doing anything manually, I'd like to do another force replication ("Get all") to se if UNM and ORC can retrieve the missing entries from UCSB.

In the process of looking at this, I saw an identifier ('rkarimi.6.3') momentarily appear as a "problem" pid, and then resolve because it was replicated to the other nodes. This reinforces my theory that it is a simple Metacat-Metacat replication issue.

Also, all the pids are ARKs from CDL.

#2 - 2012-08-17 18:58 - Ben Leinfelder

- Status changed from In Progress to Closed

This has been resolved by the epic 14-hour replication we did last night.

Files

missingORC-Identifiers.out	4.15 KB	2012-08-15		Robert Waltz
missingUNM-Identifiers.out	5.84 KB	2012-08-15		Robert Waltz